DOI: 10.1111/risa.13886

ORIGINAL ARTICLE

Information effects on lay tradeoffs between national regulatory costs and benefits

Branden B. Johnson[1] | **Adam M. Finkel**[2]

[1] Decision Research, Springfield, Oregon, USA

[2] Clinical Professor of Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor, Michigan, USA

Correspondence
Branden B. Johnson, Senior Research Scientist, Decision Research, P.O. Box 72538, Springfield, OR 97475 USA.
Email: branden@decisionresearch.org

Funding information
U.S. National Science Foundation (Division of Social and Economic Sciences: 1629287)

Abstract
A novel stated-preference "macro-risk" approach introduced to estimate the life-prolonging benefits of proposed environmental, health, and safety regulations may answer questions unasked or wrongly answered by conventional revealed-preference (e.g., "wage premiums" for high occupational risks) and stated-preference methods (e.g., willingness to pay for tiny reductions in one's own premature death risk). This new approach asks laypeople to appraise directly their preferred tradeoffs between national regulatory costs and lives prolonged nationwide (regulatory benefits). However, this method may suffer from incomplete lay understanding of national-scale consequences (e.g., billions of dollars in regulatory costs; hundreds of lives prolonged) or tradeoffs (e.g., what are lives prolonged worth?). Here we (1) tested effects of numerical contextual examples to ground each hypothetical regulatory tradeoff, and (2) explored why some people implicitly offer "implausible" values (< \$10,000 or > \$1 billion) for the social benefit of prolonging one life. In Study 1 (n = 356), after testing their separate effects, we combined three contextual-information aids: (1) comparing hypothetical regulatory costs and benefits to real-life higher and lower values; (2) reframing large numbers into smaller, more familiar terms; and (3) framing regulatory costs as having diffuse versus concentrated impacts. Information increased social benefits values on average (from \$4.5 million to \$13.8 million). Study 2 (n = 402) found that the most common explanations for "implausible" values included inattention, strong attitudes about regulation, and problems translating values into responses. We discuss implications for this novel stated-preferences method, and for comparing it to micro-risk methods.

KEYWORDS
calibration, cost-benefit analysis, national tradeoffs, stated preferences, valuation

1 | INTRODUCTION

Regulatory analysis, especially cost-benefit analysis (CBA), depends on risk estimates of the harms a regulation might prevent (e.g., lives prolonged, morbidities, environmental impacts), which are then monetized (e.g., lives prolonged multiplied by the value of a statistical life [VSL]) to allow comparison of these benefits in common units of dollars to the regulation's costs to the economy (e.g., increased prices, fewer jobs). The balance of costs and benefits for different policy options helps decide whether and how stringently to control hazards. Both conventional approaches to estimating VSL, discussed below, entail a "micro-risk" analysis of the behaviorally implied or explicitly stated value of life, based on small changes in people's personal risk of mortality or morbidity (e.g., a hypothetical reduction from 5 chances in 10,000 to 1 in 10,000).

A novel stated-preference approach that we proposed elicits lay views of acceptable tradeoffs "on behalf of the nation" between nationwide regulatory costs and benefits (Finkel & Johnson, 2018; cf. Jones-Lee et al., 1985). Arguments for such a "macro-risk" approach include that it better parallels the decisions regulators actually face, allows respondents to express their "willingness to jointly contribute" to shared-purpose risk reduction (and/or altruism), as opposed to personal benefit alone, and does not force lay decisions based on tiny, unfamiliar changes in probability of personal mortality or morbidity. However, one obvious criticism of a macro-risk approach holds that laypeople may be unable to grasp the large numbers involved in national regulatory tradeoffs (e.g., billions of dollars, hundreds of lives prolonged). We therefore pursue two questions here: 1) whether providing contextual information about costs and benefits in other realms changes the imputed estimates of the value of a life prolonged by a hypothetical regulation, and 2) why some people in macro-risk elicitations provide responses that seem to impute "implausible" values of prolonging a single life (< $10,000, > $1 billion).

2 | BACKGROUND

2.1 | Values of a statistical life

The conventional approach to regulatory cost-benefit analysis—usually requiring that benefits exceed costs to warrant adopting a given regulation—assumes benefits must be put in monetary terms for feasible comparison. For example, suppose the sole benefit of a proposed regulation is to prolong lives, so people who might die in a given year from a particular cause instead die later; for this example and current studies, we ignore other potential benefits (e.g., non-fatal diseases or injuries postponed; environmental protection). Regulators want to estimate the VSL so they can multiply it by the number of lives estimated to be prolonged each year by the policy intervention. The "statistical" portion of the term VSL refers to the usually unknown identity of those who live longer (Hammitt & Treich, 2007), and to the assumption that the VSL is useful only when no one's added risk is "large" relative to the baseline risk of death at that age (e.g., Viscusi, 2020).

VSL estimates are derived in multiple ways. Revealed preference methods entail inferring people's implicit VSL beliefs from their actual financial transactions. For example, estimating the "risk premium" in income received by people who work at jobs more dangerous than average assumes one would work in this job *only if* fully informed about its extra daily risk of death (which also presumes the employer's knowledge of these risks' magnitude, uncertainty, and cross-person variability, and disclosure of these facts to employees), *and* that one has negotiated extra compensation to fully justify that risk (which also presumes the relative power of employer and employee in such negotiation, and ability to measure compensation given that in the United States, this combines wages and benefits). By contrast, stated preferences researchers ask people directly how much their own life (usually) is worth. One prominent approach, cited earlier, presents a scenario in which one could reduce his or her risk of death from a given hazard by a small amount, and asks how much one would be willing to pay to "purchase" that risk-reducing opportunity. A related method asks whether the amount of risk reduction specified by the researcher is or is not acceptable at a specific money amount also provided by the researcher. With one such question or a series, ultimately the researcher determines the "highest" amount one is willing to pay (probably not the true maximum, for technical reasons omitted here). This willingness-to-pay (WTP) number, however derived, is multiplied by the risk reduction's magnitude, summed with others' similar results, with aggregate results possibly trimmed somehow (e.g., to remove seemingly extreme results),[1] and the average is reported as the VSL. Regulatory officials can then review both revealed- and stated-preference studies to select a VSL figure that they deem pertinent to their regulatory domain (e.g., US Environmental Protection Agency, 2016).

A huge literature covers strengths and weaknesses of CBA overall, of the VSL concept, and of revealed and stated preferences methods. Revealed preference data have historically been used more in regulatory decisions, despite the method's debatable underlying assumptions (see workplace example above) and its use of inference rather than direct questioning on the value of life. However, stated preference data have inherent advantages for regulatory decisions (e.g., Alberini, 2019), and our goal here is to help improve the quality of stated preference methods by reporting experimental results from a novel stated-preference method, which asks laypeople to evaluate which tradeoffs between US-wide regulatory costs and regulatory benefits in lives prolonged they would accept on the nation's behalf. Compared to the micro-risk approach to stated preferences, this macro-risk approach may offer its own insights for regulatory decisions, illuminate public perceptions of reasonable tradeoffs, and reveal similarities and contrasts between micro-risk and macro-risk approaches.

2.2 | Eliciting lay national tradeoffs

We introduced the concept of national-level stated-preference elicitation for regulatory analysis (Johnson & Finkel, 2016; cf. Jones-Lee et al., 1985 whose 20+ WTP questions included two national choices). Our initial focus was to understand how laypeople interpreted and responded to uncertainty in regulatory estimates of costs and benefits, part of a project on how economists versus risk analysts address uncertainty. This required identifying at the individual level which mean-preserving comparisons of uncertain outcomes people could answer meaningfully (e.g., in a nonregulatory context, people would likely be indifferent to choosing between a $100 ice cream cone, or one with uncertain cost of between $80 and $120, rejecting both choices as ludicrous; Johnson & Finkel, 2016). The need to develop personalized uncertainty ranges for gauging the acceptability of a hypothetical regulation evolved into an alternative stated-preferences method (Finkel & Johnson, 2018), involving a multiple-bounded discrete choice preference elicitation method (cf. Baik et al., 2019).

Table 1, derived from our earlier studies (Finkel &

[1] Only about 35% of 95 micro-risk stated preferences studies conducted between 1973 and 2009 reported using trims; if and how many studies trimmed but did not report those trims is unknown (Braathen et al., 2011). At least eight exclusion types were used: if uncertain of their number; failed test of understanding of probability; values above $X; protest answers; unspecified outliers; top/bottom percentile (usually 5% or 2.5%; one excluded top 15%); *N* values from high and low ends; WTP above X% of person's income.

TABLE 1 Differences between micro-risk and macro-risk stated preference studies

Attributes	Micro-risk	Macro-risk
Tradeoff perspective	Personal	National
Framing	Life first (fractional reduction in one's own mortality)	Lives first (choosing costs to justify a fixed number of national lives prolonged) OR costs first (choosing the number of lives prolonged to justify a fixed national regulatory cost)
Values used	For example, risk of death reduced from 1/10,000 to 1/100,000; dollars respondent willing to pay for this reduction in own risk	For example, costs justifying 100 lives prolonged, or lives justifying $1 billion in regulatory costs, in separate conditions (each person does one or the other)
Researcher constraint on bids	Most studies offer subjects a limited number of fixed bid values they can accept or reject.	No bid values offered; respondents can offer any value they wish
Bounds on responses	Single (final) judgment elicited; result is less reliable, does not allow respondent to show uncertainty	Double bounds (one bound marking the transition between sure support for the regulation and being unsure; the other between sure opposition to the regulation and being unsure) allow for more reliability and expression of uncertainty; geometric mean (GM) of these bounds used as imputed value
User chance to revise answers	The most common practice on this is unknown	A value is imputed (GM of provided bounds); if respondent finds this unacceptable, they can revise their bounds; if second GM rejected, they can provide an exact value
Salience	No one buys personal risk reduction from government, versus private sector; value of a statistical life is derived from judgments of the value of one's own life; excludes both altruism and shared (national) purpose implied by regulation	Costs-first frame consistent with government decisions about how best to allocate regulatory resources; either frame might include nonpaternalistic altruism component, which economists worry may involve some double-counting of benefits

Johnson, 2018; Johnson & Finkel, 2016) and our subsequent macro-risk research experience, compares several attributes of the conventional and the novel macro-risk stated preference approaches. To help inform national regulatory decisions about the VSL, the micro-risk approach aggregates personal-risk-reduction WTP of lay respondents within largely researcher-defined limits. These limits include offering WTP bids that people could only accept or reject, and framing the tradeoff as solely a judgment of lives-first (LF; i.e., fixing the amount of risk reduction, and asking about WTP for that risk reduction either personally, as in the micro-risk approach, or nationally, as in the macro-risk approach). The obverse of the LF frame is the costs-first (CF) frame, in which the amount of expenditure is fixed, and people are asked how much risk reduction they would require to make this expense reasonable (a question which micro-risk studies never ask).

Among many contrasts to the conventional stated-preference approach to VSL, the novel method elicits a national-level judgment of the tradeoff between regulatory benefits and costs, a task closer to a regulator's decision; it uses both a national-level (as opposed to individual-level, as in micro-risk) lives-first frame and a novel costs-first frame; it allows respondents to freely choose the values they offer, rather than respond yes/no to researcher-defined bids; and it asks respondents to provide bounds on their uncertainty about point estimates (e.g., between a value making the regulation acceptable and a value where one is unsure whether it is acceptable), rather than to choose a single tipping point between "definitely acceptable" and "definitely unacceptable." Other observed differences—for example, on micro-risk studies' lack of transparency and consistency on data trimming or estimates of central tendency (Finkel & Johnson, 2018)—are not intrinsic to the elicitation methods used. Some of these differences could be made smaller between the two methods. For example, a small number of micro-risk studies have elicited open bids, like the macro-risk approach.[2] For an example in the other direction, the macro-risk method could drop the CF frame and use only the LF frame. Regardless of any partial convergence, the difference between using small changes in the risk to oneself, versus average changes in the risk to everyone in the nation (and between eliciting personal costs vs. personal contributions to group costs), will persist.

The fundamental question is whether we should seek "the value of a statistical life" as the desired measure for regula-

[2] Of 95 micro-risk stated preference studies and 1140 VSL estimates 1973–2009, only 17 (17.9%) studies and 228 (20%) estimates involved open-ended elicitation (Braathen et al., 2011).

tory policy, or instead seek to estimate "the social benefit of N lives prolonged" (SBNLP; Finkel & Johnson, 2018, p. 460). Neither approach is fully satisfactory: the conventional VSL-times-fatalities measure is equivalent instead to "the average of personal valuations of N statistical lives like mine" (p. 460), more an expression of the inchoate "worth" of a statistical life than of the worth of any actual life-prolonging policy intervention, while the national stated-preferences approach might also fall short in its current incarnation by double-counting certain types of altruism, potentially biasing benefits estimates upwards (Finkel & Johnson, 2018, p. 462).

Previously, we used the geometric mean of the two bounds on acceptable tradeoffs provided by respondents to define what we call here the SB1LP* value (i.e., the social benefit of one life prolonged; the asterisk indicates further adjustments—for example, to avoid double-counting of non-paternalistic altruism—have not yet been made to the method to achieve true SB1LP, although these adjustments are now being estimated; Finkel & Johnson, 2018). We found in our earlier sample (Section 3), after combining responses from people confronted with either a hypothetical regulation involving a carcinogen or one to reduce traffic accidents:

- a broad, right-skewed distribution of SB1LP* values, probably reflecting the unconstrained elicitation here, contrasting with largely researcher-defined bids in micro-risk studies (Finkel & Johnson, 2018);
- means were lower in the LF (n = 340) than the CF (n = 393) frame, regardless of trims or control variables used (Finkel & Johnson, 2018; Table 1);
- those means were higher in the full sample ($31 million full, $28 million LF, $35 million CF) than the current agency-used VSLs, but similar when we trimmed the top and bottom 5% ($5.5 million, $6.5 million, $4.3 million; that study did not identify or trim implausible values); and
- linear regression analysis found that income decreased imputed SB1LP* values; a limited set of other potential predictors (age, risk type, exposure to information about agency-used VSLs, ratio of upper to lower bound, expecting regulators to underestimate cost or accurately estimate deaths postponed) were nonsignificant at $p < .05$, implying generalizability of these imputed values to the population at large.

2.3 | Current studies

While intriguing, the national-level elicitation results to date come from a single dataset. They need confirmation or refutation, a major goal of studies reported here. Specifically, we wondered whether our earlier (Finkel & Johnson, 2018) findings generalized to a different opportunity sample regarding (1) the difference in SB1LP* imputed values between CF versus LF frames; (2) comparisons to agency-used VSLs; and (3) predictors of SB1LP* values, using a broader set of such variables.

We also had other goals. First, we wished to probe the degree and causes of unusually high and low SB1LP* values. Earlier we (Finkel & Johnson, 2018) reported one trim excluding people whose imputed SB1LP* value was ≤ $100 or ≥ $1 billion, mostly on the low side, comprising 13.1% of the full sample of 733, 2.5% of the CF frame, and 25.3% of the LF frame. What we then labeled "outlying values" we have now chosen to call "implausible" monetary values per human life prolonged. Given no objective definition of a plausible VSL, we decided to omit as "implausible" imputed SB1LP* values of < $10,000 or > $1 billion. We felt the earlier $100 lower bound (Finkel & Johnson, 2018) excluded too few implausible values, and we note that $10,000 is roughly 1/1000 of the upper end of (~ $10 million) VSL values used by US federal agencies (Robinson et al., 2021). Also, if a statistical life was only "worth" $10,000, half of the $6.6 trillion US federal budget for 2020 could "save" everyone. The $1 billion value (100 times the high end of current agency values; 100,000 times our low "plausible" value) would be able to "save" only 6000 Americans annually (a risk of 2/100,000 per year) with the entire federal budget used up. Obviously, the definitional choice of what extremely low or high responses are "implausible" is a worthy subject for continued debate. Here we sought to clarify the proportion of people who offer implausible responses, and to explore why such responses occur (including the substantial difference in the proportion of implausible values between the LF and CF frames in Finkel & Johnson, 2018, data).[3]

Second, we needed to decide how best to trim outliers from the data, as this can determine such issues as how the SB1LP* values compare to agency-used VSLs; Finkel and Johnson (2018) had shown three trims of macro-risk data without determining a preferred option. After considering our prior results, various trims in prior micro-risk VSL elicitations (when infrequently specified), and other options (e.g., exclude speedy responses), we adopted a triple trim, inspired by multiple-step trims such as those by Carson and Mitchell (2000) and Andersson (2006): (1) we included only respondents who got at least one of two attention checks correct (i.e., buried in a lengthy question were instructions on how to answer it, which the inattentive would overlook); (2) included only those whose imputed SB1LP* value is "plausible" (i.e., $10,000–$1 billion, inclusive); and (3) used only the values from the middle 90% of the resulting distribution, that is, with the top 5% and bottom 5% trimmed. This approach minimized inclusion of values produced by the inattentive, uncomprehending, or protesting respondent, while developing an expected-value estimate that allows influence by large values (as the median does not) without allowing the mean (the preferred estimate for benefit-cost analysis [e.g., Alolayan, 2012]) to be distorted by outsized right-tail outliers. Using a (trimmed) expected-value estimate is also consistent conceptually with paying for a regulatory program if it were based on willingness to pay (e.g., the government would

[3] The frequency of implausible values in micro-risk studies is unknown, as (1) very few investigators ever report raw data or probability density function percentiles, and (2) by definition implausible values are nearly impossible when the researcher restricts the value of bids which you can only accept or reject. Also see footnote 1, and Finkel and Johnson (2018, p. 471) on limited distributional data from micro-risk studies (e.g., Braathen et al., 2011).

collect N * mean, not N * median). By proposing this as a standard approach, we provide scholars with a concrete proposal to debate, and to inform discussions of whether similar standardization would benefit micro-risk studies.

Third, reviewers of our current U.S. National Science Foundation (NSF) project narrative had expressed concern about citizens' potential inexperience with large numbers (e.g., costs in millions or billions, and lives prolonged across the nation of 100 or more), whose effects on results we probe here. If laypeople can readily grasp the relative value of different magnitudes, particularly for large regulatory costs, it is more plausible that SB1LP*s derived from their responses are valid and reliable (Finkel & Johnson, 2018). A similar question has long been raised about the degree of, and how to improve, public understanding of the small probabilities inherent in micro-risk elicitations (e.g., Knoblauch et al., 2017; Visschers et al., 2009; Weinstein et al., 1994). Concerns about comprehension of small probabilities have sometimes been addressed in conventional VSL studies through such means as grids, risk ladders, frequency statements, and other aids (Cropper et al., 2011). The challenge in either setting is how to define, and measure, that people "grasp" large numbers or tiny probabilities. We proposed to test the effects of providing information that put our hypothetical regulatory benefits and costs into the context of various real expenditures and life-prolonging activities that people pursue, collectively or at the household level. We had no prior expectations as to whether better calibration would decrease or increase responses relating to the value of life. On the one hand, the high proportion of low implausible values in the LF frame (Finkel & Johnson, 2018) might favor a sense that providing contextual information would increase SB1LP* means. Yet better calibration might reduce very high values instead. Other outcomes are also possible and difficult to exclude *a priori*: for example, laypeople may be well-calibrated for national regulatory numbers already, so that low values are a feature rather than a bug, or that laypeople are simply insensitive to contextual information.

In summary, we posited the following hypotheses (H) and research questions (RQ):

H1: *Imputed SB1LP* mean values will be lower for lives-first (LF) than costs-first (CF) frames*. This was the finding of earlier research, although its source was unclear; we speculated that it might reflect enhanced concern about financial impacts to households or the economy when people have to come up with dollar amounts themselves, compared to when the nation seems to have a fixed amount to "spend" (Finkel & Johnson, 2018).

H2: *Imputed and trimmed SB1LP* mean values will be close to current US agency-used VSLs*. This was our earlier finding (Finkel & Johnson, 2018), so we are extrapolating that to our newly proposed trim. However, we are assuming also that both the micro-risk and macro-risk stated-preference approaches faithfully (if incompletely; see earlier "fundamental question" discussion) elicit what people believe about life-cost tradeoffs, and thus should yield roughly similar results.

H3: *Implausible imputed SB1LP* values will be more common in the LF than the CF frame*. This also was an earlier finding (Finkel & Johnson, 2018), which if the financial concern referred to above is really a factor, may depress respondents' suggested bounds on reasonable tradeoffs much more in the LF frame than in the CF frame, whose starting value (e.g., $1 billion in regulatory costs) they may treat as an already fixed "expense" (see Johnson & Finkel, 2016 on respondents tending to treat regulatory costs as part of the federal budget, as with construction projects and other expenditures).

RQ1: *What demographic, political, or regulatory variables affect imputed SB1LP* values?* Prior results (Finkel & Johnson, 2018) suggested income might be the only factor among several tested that differentiated SB1LP* values, but this warranted further examination.

RQ2: *How does exposure to contextual information affect imputed SB1LP* values?* We assume that generally responses will become more valid the more information people get, subject to constraints of numeracy and cognitive involvement. Answers to this question will help determine whether people can calibrate the large values in macro-risk studies.

RQ3: *Why do many lay responses to macro-risk questions yield implausible imputed SB1LP* values?* Assuming implausible values are a feature of the macro-risk process (e.g., allowing open bids), understanding their sources can improve calibration of respondents in future macro-risk studies.

Before we report results of our new studies, we report reanalyzed data from Finkel and Johnson (2018). Although our hypotheses and research questions partly stemmed from this prior study of ours, our currently proposed trims were not applied therein. Further, the two hypothetical regulations that we randomly assigned to respondents in that earlier sample were combined in that article's SBNLP estimates (Finkel & Johnson, 2018), possibly obscuring differences between responses to the two regulations.

3 | REANALYSIS OF FINKEL AND JOHNSON'S DATA

Because methods and results have been published elsewhere (Johnson & Finkel, 2016; Finkel & Johnson, 2018), we focus here mainly on those specific to comparing results to those of our new studies.

3.1 | Methods

3.1.1 | Sampling

US residents 18-plus years old and proficient in English were randomly recruited by Johnson and Finkel (2016) in June–July 2012 from the diverse but nonrepresentative Decision Research national online panel; members who answer surveys were reimbursed at the rate of $15 per hour.

3.1.2 | Instrument

As these earlier data were not originally collected to develop a macro-risk approach to estimate imputed SB1LP* values (see Johnson & Finkel, 2016, for details), and its elicitation method was almost identical to that used in our Study 1 detailed later, here we describe only the core approach. Laypeople were shown a hypothetical proposed regulation; half were told the agency estimated its annual regulatory benefits as 1000 lives prolonged nationwide (LF frame), half that it estimated annual regulatory costs at $1 billion (CF frame). The elicitation process focused on national rather than personal impacts *and* elicited double bounds rather than accept or decline bids offered by researchers who elicit single point estimates, as in the micro-risk tradition. The two frames respectively began an iterative process to elicit a value of national costs low enough, or national number of lives prolonged high enough, for the respondent to "definitely support" the regulation. In each frame, the respondent was then asked to provide another value low (high) enough for the respondent to "definitely oppose" the regulation. Both prompts were described as "if this answer was any higher [or lower, depending on the frame], you should feel 'unsure' about whether to support or oppose the regulation at that level of cost [or lives prolonged]." An on-screen slider revealed the value people were currently considering, with the LF version also providing per US household and per capita feedback on the current national regulatory cost value being considered. The geometric mean of the two bounds was assumed to be a reasonable proxy for the tipping point (Finkel & Johnson, 2018).

3.1.3 | Analyses

As noted earlier, we reanalyzed the data collected by Johnson and Finkel in 2012 (with imputed SB1LP values published in Finkel & Johnson, 2018) for this article by using a two-trim method (middle-90% of plausible SB1LP*; the third trim requires an attention check not present in the prior instrument) to ease comparison of those results to our new data reported in Sections 4–5, along with full-sample (untrimmed) results. Our emphasis here is on (trimmed) mean results, but for each analysis we also report the 25th, 50th (median), and 75th percentiles of the distribution.

3.2 | Results

3.2.1 | Sample

Some 744 responses (after the authors removed 52 people for incomplete, incoherent, or other unusable answers) came from a sample slightly more female; more educated, and higher income; and about the same degree of non-Hispanic White ethnicity as in the US population (Johnson & Finkel, 2016).

3.2.2 | Imputed SB1LP* Values

The top half of Table 2 shows imputed SB1LP* values for both the CF and LF frames using full and two-trim (middle 90th-percentile plausible) samples, first for the US Environmental Protection Agency (carcinogen) and second for the National Highway Traffic Safety Administration (NHTSA; traffic safety) hypothetical regulations. H1 posited that imputed SB1LP* means will be lower for LF than CF frames, which is the case here except for the trimmed NHTSA values. H2 posited that imputed SB1LP* means will be closer to current US agency-used VSLs after trimming, also the case here, with the four trimmed means ranging from $6.9–$10.7 million. H3 posited that implausible imputed SB1LP* values will be more common in the LF than the CF frame. Based on these samples trimmed to the middle 90th percentile of plausible values, proportions of implausible values were 44.5% and 46.2% for the LF carcinogen and traffic regulations, versus 20.9% and 15.3% for their CF equivalents, consistent with the hypothesis.

4 | STUDY 1: CONTEXTUAL INFORMATION

The first new study we report here tests the effect of contextual information on SB1LP* values within-person.[4]

4.1 | Methods

4.1.1 | Sampling

US residents 18-plus years old proficient in English were randomly recruited May 21–22, 2019 from the Prolific online panel, whose members receive small fees for answering surveys.

4.1.2 | Instrument

After screening questions (minimum age, US residence, having time available now to complete the survey in one

[4] A prior study, designed similarly to the one reported here but whose responses raised our concerns about quality (we changed sample vendors as a result), is reported in the Supporting Information.

TABLE 2 Selected distributions of SB1LP* in the first tradeoff: Reanalyzed Finkel and Johnson (2018) data and current studies 1–2

Percentiles	**Costs-first frame**			**Lives-first frame**		
(Finkel & Johnson, 2018)	**Full**	**Two-trim**	**Three-trim**	**Full**	**Two-trim**	**Three-trim**
Carcinogen	(n = 191)	(n = 151)		(n = 182)	(n = 101)	
25th	$63,241	$166,725	NA	$316	$187,531	NA
50th	$1,000,000	$1,291,219	NA	$97,421	$948,418	NA
75th	$10,000,000	$8,165,824	NA	$1,975,174	$4,059,386	NA
Mean	$39,330,512	$10,691,695	NA	$36,084,077	$6,851,568	NA
Traffic safety	(n = 202)	(n = 171)		(n = 158)	(n = 85)	
25th	$127,246	$223,357	NA	$12	$104,824	NA
50th	$805,455	$1,153,453	NA	$42,896	$805,378	NA
75th	$4,477,133	$3,647,539	NA	$961,314	$5,286,251	NA
Mean	$30,459,841	$7,643,874	NA	$18,651,066	$8,390,508	NA
Study 1						
Preinformation	(n = 129)	(n = 88)	(n = 73)	(n = 162)	(n = 40)	(n = 33)
25th	$9,428	$41,181	$48,188	$1	$52,290	$40,307
50th	$129,032	$316,456	$289,017	$6	$223,607	$223,607
75th	$1,408,451	$1,481,722	$1,428,863	$14,827	$1,731,391	$1,327,591
Mean	$3,279,539	$1,965,223	$1,661,773	$6,308,237	$4,807,260	$4,509,731
Post-information	(n = 123)	(n = 102)	(n = 80)	(n = 168)	(n = 64)	(n = 50)
25th	$22,361	$59,693	$69,800	$1	$6,397	$10,282
50th	$316,456	$459,063	$459,063	$4	$179,057	$227,940
75th	$2,564,103	$2,860,360	$2,518,315	$44,528	$3,162,278	$6,922,866
Mean	$14,635,996	$4,903,925	$4,369,346	$13,311,000,000	$12,054,665	$13,799,419
Study 2	(n = 202)	(n = 149)	(n = 117)	(n = 200)	(n = 124)	(n = 102)
25th	$55,801	$316,212	$316,180	$483	$223,607	$254,333
50th	$999,500	$1,412,801	$1,154,585	$269,218	$2,737,089	$3,162,278
75th	$7,114,618	$6,351,907	$6,044,910	$13,668,464	$23,961,343	$22,894,234
Mean	$130,766,927	$9,824,358	$8,672,012	$397,206,463	$21,882,061	$19,287,611

sitting), we included an oath to provide accurate answers; this truth-telling-commitment device outperformed other options in a second-price auction for nonmarket goods (Jacquemet et al., 2013). We followed with background information on regulatory benefits (reducing mortality) and costs, on the nature and unavoidability of tradeoffs in life and policy, and on the respondent's role to decide "on *behalf of the nation* rather than just yourself, whether a specific regulation's total national benefits are worth the total national costs" (emphasis in original). We included two separate attention check questions (e.g., after a long introduction to a question about personal preference, tell them to give a specific answer regardless of their actual preference), before and after the first tradeoff described below, as inattention might explain implausible SB1LP* values (Maniaci & Rogge, 2014); multiple checks out-perform single ones (Berinsky et al., 2014).

Then the single hypothetical regulatory scenario posited that NHTSA was considering requiring that all old and new cars include a collision warning device, to alert the driver so as to effectively reduce the number of fatalities to passengers or others (e.g., pedestrians) outside the car. Each respondent was randomly assigned to either a CF frame, in which NHTSA had estimated the regulation would cost the nation exactly $1 billion ($1,000,000,000) annually but had not determined the number of lives prolonged nationally, or a LF frame in which NHTSA had estimated the regulation would benefit the nation by prolonging 100 lives per year, but had not determined its cost. CF respondents were asked to indicate upper and lower bounds for the number of lives prolonged that would make this annual regulatory cost either definitely accepted or definitely rejected "on behalf of the nation"; LF respondents were asked to report upper and lower regulatory cost values that would make the 100 lives prolonged each year definitely rejected or definitely accepted. Between these bounds was the region where respondents indicated they were unsure about their support or rejection.[5] After confirming these two bounds and respondents'

[5] In this study only (versus Finkel & Johnson, 2018, and our other macro-risk studies), we elicited two additional bounds (between "slightly unsure" and "really unsure"). As comparing geometric means of the two inner versus the two outer bounds yielded no substantive differences in imputed SB1LP* values, we report results from the two outer bounds, as discussed in the text.

TABLE 3 Study 1, information treatment

Costs-first	Lives-first
To put this choice into context, $1 billion per year is • half of the $2 billion that heart transplants cost Americans each year (about 2,500 such operations are performed each year, and about half of the recipients live at least 10 additional years). • 100 times the $10 million cost per year of ejector seats to allow pilots an emergency bail-out from one kind of jet fighter plane (about two pilots die per year from failure to eject) • equivalent to about $3 for every person, or about $10 for the average-sized household, in the U.S., paid each year; the U.S. national economy is $19 trillion per year, 19,000 times bigger than this • equivalent to everyone paying a little less than one additional penny per gallon of gasoline, which almost every household buys often • equivalent to a $105 rise in the price of a refrigerator ($905 versus $800 for a low-cost bottom-freezer refrigerator), which households replace about every seven years To put the numbers you might offer for lives-prolonged into context, each 100 lives prolonged nationally each year is equivalent to • entirely preventing a year's deaths from murders committed via poison in the U.S. (about 90 per year) • reducing by about 20% the number of annual deaths from gun-related accidents • the equivalent of a 1 in 43 (about 2.3%) chance that at least one of your friends and acquaintances—if you had 1,000 of them—would be among the 100 Americans who would have their lives prolonged by this regulation. • About 2.6 million Americans die (of any cause) each year.	To put this choice into context, 100 lives prolonged nationally each year is • 4%, or 1/25, of the annual number of lives prolonged in the U.S. each year thanks to heart transplants (these operations cost in total about $2 billion per year) • 50 times the roughly two lives prolonged per year by ejector seats to allow pilots an emergency bail-out from one kind of jet fighter plane (at a total cost of about $10 million) • equivalent to entirely preventing yearly deaths from murders committed via poison in the U.S. (about 90 per year) • equivalent to reducing by about 20% annual deaths from gun-related accidents • the equivalent of a 1 in 43 chance that at least one of your friends/acquaintances—if you had 1,000 of them—would be among the 100 Americans who would have their lives prolonged by this regulation during their lifetimes. • About 2.6 million Americans die (of any cause) each year. To put the numbers you might offer for cost into context, each $1 billion per year is equivalent to • about $3 for every person, or about $10 for the average-sized household, in the U.S., paid each year; the national economy is $19 trillion per year, 19,000 times bigger than this • everyone paying a little less than one more penny per gallon of gasoline, which almost every household buys often • a $105 rise in the price of a refrigerator ($905 versus $800 for a low-cost bottom-freezer refrigerator), which households replace about every seven years

Note: Shading indicates conventional and novel perspective clauses for CF frame (LF is identical, but left unshaded).

confidence in them, we reported to respondents their bounds' geometric mean (GM), asking whether they viewed their GM as a reasonable tradeoff (the GM being more appropriate than an arithmetic mean; see Finkel & Johnson, 2018, note 56). If a respondent did not accept that value, they were invited to revise their bounds to produce a value more acceptable; if they did not accept the resulting new GM, we invited them to provide an acceptable point estimate of any size.

After the second attention check, the respondent was told that this elicitation would be repeated with added contextual information. Table 3 presents the information treatments for both frames. These each began with their own frame's information, and only then presented the information relevant to the other side of the ledger. We combined three types of contextual information in this test of information effects on SB1LP* values, based on prior results (footnote 4). The first type provides actual regulatory or other risk-related costs and benefits higher and lower than the CF ($1 billion) or LF (100 lives) starting points (e.g., heart transplants; ejector seats); these analogies aimed to increase laypeople's sense of how big or small were values they juggled. The second type involves perspective clauses, which re-express large numbers in familiar terms. Using ratios, ranks, and unit changes, experiments (combined $n > 3200$) found that adding such clauses greatly improved subjects' ability to recall measurements, estimate other quantities, and detect errors in manipulated measurements (Barrio et al., 2015); our examples include dollars per capita or household, average risk reductions, and the magnitude of the US economy or nationwide deaths per year. The third type was a novel perspective clause highlighting frequent versus occasional costs, using gasoline and refrigerator prices. Both perspective clause types are shaded in the CF side of the table to make them more visible.

After the second tradeoff, respondents were asked to explain any differences in their two accepted values (with and without the new information): "Please explain what rethinking of values, information, or other factors affected your thinking." The instrument ended with demographic and other measures to help put main results into context. The latter included subjective (reported comfort with numbers; Fagerlin et al., 2007) and objective numeracy (five probability-focused multiple-choice computational questions), perceived accuracy of agency estimates of regulatory costs and benefits, trust in environmental, health, and safety agencies to make appropriate regulatory decisions, regulatory preferences (support for government regulating businesses' and/or individuals' behavior, or neither), and regulatory trend preference (the current administration should regulate more than, less than, or the same as the last two US administrations).

4.1.3 | Analyses

From the final value provided in each tradeoff (first GM, second GM, or exact number), we divided $1,000,000,000 by the GM (in the CF frame) or divided the GM by 100 (LF) to

TABLE 4 Study samples

		Study 1	Study 2	US Adults
Female		66.6%	51.5%	51.3%
Age	M (SD)	50.5 (13.5)	34.3 (13.3)	38.2 (includes children)
	Median	51.0	31.0	NA
Education	≤ high school degree	16.3%	13.7%	38.6% (25+ years old)
	≥ college degree	53.9%	50.3%	32.6% (25+ years old)
Non-Hispanic white ethnicity		84.2%	NA	72.2% (includes children)
Political ideology	Liberal	37.2%	NA	36%
	Conservative	33.0%	NA	27%
Political partisanship	Democrat	32.0%	NA	28%
	Republican	34.0%	NA	27%
Employed full-time		40.8%	NA	NA
Household income: median range (% of sample/population)		\$30,000–\$59,999 (27.4%)	\$30,000–\$59,999 (29.6%)	\$50,000–\$74,999 (17.4%)
Wealth: median range (% of sample)		\$20,000–\$74,999 (13.8%)	Zero–\$19,999 (32.8%)	NA

Note: NA = not available. Study 2 omitted several demographic questions, substituting questions about "implausible" responses (see text). US data are from 2018 ACS 1-year estimates of the US Census Bureau, other than political ideology (from a May 3–7, 2017 representative survey Pew Research Center, 2017) and political partisanship (from an April 25–May 1, 2018 representative survey Pew Research Center, 2018).

yield an imputed SB1LP* value. Our primary analyses here were descriptive or correlational, given the highly skewed nature of this bid-free distribution, and the results shown below.

4.2 | Results

4.2.1 | Sample

Respondents as a group (n = 356) were more female, white, and educated than US adults, with similar political ideology and party affiliation to US adults generally (Table 4). Median completion time was 21 min. Random assignment yielded 58.1% LF respondents. Three-quarters (77.6%) got one or both attention screeners correct.

4.2.2 | SB1LP* values

Some 291 respondents (81.7%) in the sample could define an acceptable GM/life value. Plausible values (\$10,000–\$1 billion) overall comprised 75.2% (CF) and 26.5% (LF) of respondents, consistent with H3 that the LF frame yields more implausible values. CF implausible values were 31.8% (preinformation) and 17.1% (postinformation), and LF implausible values 75.7% and 61.1%, suggesting that providing information did reduce implausible-value frequency.

Table 2 (middle section) shows the same distribution metrics as reported earlier for Finkel and Johnson (2018) data for imputed SB1LP* values both before and after exposure to contextual information, adding our recommended three-trim including an attention check criterion.[6] Inconsistent with H1, imputed SB1LP* mean values were higher for LF than CF frames in all six contrasts. Results for H2 were mixed: imputed SB1LP* mean values were close to current US agency-used VSLs after trimming in the postinformation condition, but slightly above that range in the LF condition, while all results in the preinformation conditions fell within that agency range. If we compare the similarly trimmed traffic safety means from Finkel and Johnson (2018) to these results, all but one (LF postinformation) of the four contrasts fell within the current agency-used VSL range; CF results were larger in the earlier study, but its LF mean was bracketed by those found here.

4.2.3 | Predictors of SB1LP* values

Our first research question was which (if any) demographic, numeracy, or attitudinal variables affect imputed SB1LP* values. Given the large proportion of implausible values, our trimmed samples (particularly for LF framing) were too small to justify regression analysis for even a few predictor variables, much less the 13 we considered here. Thus we used correlational analysis to identify a possible subset of such variables worth using with the full sample. We report these in Table 5, but note that had we applied a correction factor for the multiple analyses here, no remaining correlation would be statistically significant (Glickman et al., 2014). Here we began with Ferguson's (2009) recommended minimal practi-

[6] LF full samples were larger after the information was provided than before, because fewer people rejected the geometric mean of their two bounds without substituting an exact figure.

TABLE 5 Study 1, Full and trimmed sample correlations with imputed SB1LP* values

	Pre-information				**Post-information**			
Demographics	**CF full ($n = 129$)**	**CF three-trim ($n = 73$)**	**LF full ($n = 161–162$)**	**LF three-trim ($n = 33$)**	**CF full ($n = 122–123$)**	**CF three-trim ($n = 80$)**	**LF full ($n = 167–168$)**	**LF three-trim ($n = 50$)**
Gender (female)	0.102	−0.010	0.038	−0.208	0.085	0.008	0.053	0.037
Age	−0.175*	0.022	0.035	0.101	−0.147	−0.088	0.004	0.239†
Education	0.059	0.082	0.043	0.100	0.043	0.045	−0.022	0.101
Income	0.097	−0.039	0.078	−0.012	0.070	0.031	0.031	−0.116
Wealth	0.059	0.071	−0.047	0.072	−0.083	0.050	−0.105	0.111
Numeracy								
Objective	0.132	0.191	−0.027	0.102	−0.040	0.021	−0.040	0.034
Subjective	−0.072	−0.088	−0.071	−0.080	−0.014	0.018	0.074	−0.073
Government-related Beliefs								
Democrat	0.137	0.011	−0.020	0.215	0.140	0.065	−0.055	0.038
Republican	−0.024	−0.102	−0.072	−0.092	−0.089	−0.093	0.119	−0.164
Conservative ideology	−0.181*	−0.130	−0.112	−0.090	−0.127	−0.118	0.064	−0.258†
Regulate business	−0.047	0.066	−0.095	−0.154	0.113	−0.061	−0.070	−0.012
CF accuracy	−0.086	0.007	NA	NA	0.057	−0.233*	NA	NA
LF accuracy	NA	NA	0.039	0.270	NA	NA	0.110	0.319*

Notes: three-trim: restricted to SB1LP* values for people who were correct on at least one of two attention checks, and were in the middle 90th percentile of plausible ($10,000–$1 billion) values. The CF accuracy question (is the agency accurate in estimating these costs?) is applied only in the CF correlation columns; the LF accuracy question (is the agency accurate in estimating the lives prolonged?) is applied only in the LF correlation columns.

†$p < 0.10$

*$p < 0.05$.

cal effect criterion (RMPE; $r \geq 0.2$), which he deemed more reliable and consistent with other effect sizes than Cohen's (1992) criterion for small effects of $r = 0.1$, complemented by review of top predictors. Predictors included demographics (gender, age, education, and income and wealth as complementary ways to assess ability to pay); objective and subjective numeracy (respectively the ability to use, and comfort with using, numbers, potentially relevant to the calibration issue of grasp of large numbers); and government-related variables (political partisanship, political ideology, preference for regulating business, and belief that NHTSA would provide accurate estimates of regulatory costs or benefits).

Of 96 correlations in Table 5, only seven exceeded Ferguson's (2009) RMPE criterion. Belief in accurate lives-prolonged agency estimates and being a Democrat were associated with higher SB1LP* values, and female gender with lower ones, in the preinformation LF trim; belief in accurate cost estimates was associated with lower SB1LP* values in the postinformation CF trim; and life-estimate accuracy and age were associated with higher, and conservative political ideology with lower, SB1LP* values in the postinformation LF trim. Adding 24 items meeting Cohen's (1992) small-effect criterion does not change the pattern of government-belief predictors having roughly double the number of "effective" correlations as demographics. The LF frame dominated RMPE-qualified associations, although using the Cohen criterion reduced the gap (CF 14 versus LF 18, of 48 correlations each). Conservative ideology was the most consistently "effective" predictor (six of eight comparisons exceed Cohen's small effect criterion), with a negative association with SB1LP* values. This is consistent with low values being protest votes, whether against regulations in general or this hypothetical regulation specifically. The results do not justify multiple regression analysis.

4.2.4 | Contextual information effects

For our second research question, Table 2 shows mean SB1LP* values increased from pre- to post-information tradeoffs: for CF, by factors of 4.5-fold (full sample) and 2.6-fold (three-trim), and for LF, by 2110-fold and 3.1-fold. Thus overall contextual information increased SB1LP* in all conditions, but particularly for the LF frame, apparently (see above) via elimination of many implausibly low values, which we plausibly interpret as a signal of better calibration.

However, information effects may differ at the individual level from overall results. Table 6 shows such differences, using the ratio of post- to pre-information SB1LP* values for each person, for both the full samples per frame, and for the trimmed samples (here we show only changes based on pre-information trims; post-information had differing proportions of plausible SB1LP* values, but changes' proportions were very similar—see first author for details). The proportion whose values went up after information exposure greatly exceeded the proportion whose values went down, in all four overall conditions. Major change (defined as ratios of ≤ 0.2 for decreases and ≥ 5.0 for increases) was again dominated by individuals whose SB1LP* values increased after contextual information exposure. Effects across frames—the ratio of higher to lower proportions—were higher for CF in both full-sample conditions (2.09–1.47 overall, 2.48–1.54 major), but LF dominated with trimming (overall 2.88–2.47 pre-information; major 2.01–1.26 pre-information). Thus the overall increase in SB1LP* means observed earlier was due to substantially more people increasing their values after information exposure.

We applied binary logistic regression to the full sample to assess potential predictors of increases versus decreases in SB1LP* values following contextual information exposure. Predictors tested included gender, age, education, income, objective and subjective numeracy, and political ideology. In the CF frame ($n = 105$, up $= 71$, down $= 34$) this regression model with seven predictors classified 76.2% correctly, and being female was the only predictor that significantly affected (in this case, decreased) values controlling for the other variables (OR = 0.275, 95% CI 0.094, 0.805, $p = 0.018$); the model was significant overall ($X^2 = 16.444$, $df = 8$, $p = 0.036$), with modest pseudo-R^2 values (Cox & Snell = 0.10, Nagelkerke = 0.14). In the LF frame ($n = 110$, up $= 66$, down $= 44$) this model classified 64.5% correctly, and being highly educated was the only significant influence, with this group more likely to decrease values (OR = 0.632, 95% CI 0.412, 0.969, $p = 0.035$); the model was marginally significant overall ($X^2 = 13.705$, $df = 8$, $p = 0.090$), with modest pseudo-R^2 values (Cox & Snell = 0.07, Nagelkerke = 0.10). Results should be treated cautiously, given the small number of observations, but it appears that even if the CF frame has a slightly stronger effect, there is no strong or consistent association with predictors (e.g., Ferguson's, [2009] criterion for a recommended minimum practical effect of $R^2 = 0.04$ is met in both cases, but not his parallel criterion for OR = 2.0 [0.5 for decreases], although he said the OR criteria should be used with caution).

We also examined respondents' open-ended explanations of changes in their pre- versus post-information imputed SB1LP* values, focusing particularly on their mention of the contextual information. Some people in each frame cited specific examples from our contextual information, from both the higher/lower and perspective clause examples, but no one specifically cited the frequent/occasional purchase examples. In the CF frame 10 people gave specific, sometimes multiple, examples: per person or household costs (5), pilots (4: e.g., "The dollar amount to save lives of fighter pilots really hit home. I undervalued lives immensely initially"), transplants (2), average annual deaths (2), firearms (1), and cost relative to the economy (1). Some 24 other CF respondents named the contextual information more generally (e.g., "other statistics put it in better perspective"), and five possibly referred to this (e.g., "after reading more . . . not as sure on regulation support"). Of 114 CF responses, these 39 references to the

TABLE 6 Study 1, effects of contextual information at the individual level on SB1LP* values

	Full sample			Trimmed (preinformation)		
	No change	Lower	Higher	No change	Lower	Higher
Overall						
LF	17.8%	33.3%	48.9%	0%	25.8%	74.2%
CF	7.9%	29.8%	62.3%	9.2%	26.2%	64.6%
Major change						
LF	NA	28.9%	44.4%	NA	16.1%	32.3%
CF	NA	21.9%	54.4%	NA	18.4%	23.1%

Note: Major change = ratios of ≤ 0.2 for decreases and ≥ 5.0 for increases.

information comprised 34.2% (52.7% of 74 when excluding nonresponsive and "don't know" answers). On the LF side, eight people gave specifics—per person/household (6), equivalent lives prolonged (1), probability for social network (1)—26 cited the information generally, and three were possible references. Of 120 LF responses, these 37 references to the information comprised 30.8% (49.3% of 75 when excluding nonresponsive and "don't know" answers). In short, many of our respondents were conscious of the information we provided.

5 | STUDY 2: EXPLORING "IMPLAUSIBLE" SB1LP* VALUES

Our next study explored reasons for "implausible" values, to further illuminate the viability of eliciting lay national-level tradeoffs of regulatory benefits and costs.

5.1 | Methods

5.1.1 | Sampling

A sample of Americans 18+ fluent in English was recruited from the online panel Prolific 18–19 December 2019.

5.1.2 | Instrument

We simplified the Study 1 instrument to elicit only one tradeoff, to seek only two bounds, to include the contextual information before the tradeoff, and to remove several potential predictors (Table 3), in order to focus attention on self-reported reasons for implausible results. After people accepted or rejected the geometric mean, people providing imputed SB1LP* < $10,000 or > $1 billion (implausibles) were asked whether this value still seemed reasonable to them as a household share of regulatory costs to prolong one statistical life somewhere in the nation (e.g., "it is worth spending no more than" or "no less than" so much per household).

If a subject still held onto an implausible value, they were asked why, first in an open-ended format, followed by a multiple-choice approach derived from prior studies' qualitative responses and researcher suggestions. Some questions were directed at low SB1LP* ("I am anti-regulation in general," "I am opposed to this regulation in particular," "People die from a lot of different things, so I don't think we should spend too much on any one thing"), while others were aimed at high SB1LP* responders ("I am pro-regulation in general," "I am very supportive of this regulation in particular," "If we can prevent people from dying from a particular cause, I think we should spend whatever is necessary to prevent those deaths"). Still other questions were targeted at all respondents ("I provided the first answer that came to mind," "I did not understand what was being asked," "I don't care about the question or this issue," "Life is infinitely precious and cannot be equated to a dollar amount," "Other (explain).")

People who accepted an implausible value but rejected it after seeing the per-household number were then asked to explain, with choices "The number provided implies [X lives or Y dollars per life prolonged], which is too high or low," "The number provided is in my 'uncertain' area of support," or "Other (explain)." Then an invitation to reset their bounds instructed that those bounds should be higher or lower than original bounds (varying by frame) if they thought the value too low, and vice versa if they thought it too high. They then repeated the tradeoff exercise before seeing the closing items.

5.2 | Results

5.2.1 | Sample

Of the 402 respondents, 75.9% correctly answered both attention screeners. Table 3 shows this sample was more educated and younger, but similar in terms of male/female proportions, as US adults overall. Median completion time was 17 min.

5.2.2 | Imputed SB1LP* values

Some 73.6% of respondents (CF 80.7%; LF 66.5%) had imputed plausible SB1LP* values ($10,000–$1 billion); the implausible proportion was higher for the LF than CF frame, consistent with H3. The lower proportion of implausible responses among LF subjects here compared to our prior studies may stem from exposing everyone to contextual information from Study 1, plus variability in cross-study responses when laypeople can offer any bounds they wish.

Table 2 (bottom) shows that in this study CF means were less than LF equivalents, again contrary to H1; trimmed means were close to agency-used VSLs, consistent with H2, with CF trimmed means at the upper end of the agency range and LF trimmed means about twice that high end. The pooled (CF and LF responses together) SB1LP* mean was $14,210,826, or roughly twice as much as the postinformation mean in Study 1, and about 50% higher than the typical VSL value US federal agencies currently use. Possibly this earlier result was driven by anchoring on the lower preinformation response, pulling responses down relative to the condition in this study where contextual information was presented without a prior tradeoff. Bivariate correlations of the pre- versus postinformation values in Study 1 were so low (CF: $r = 0.089$, $p = 0.344$, $n = 114$; LF: $r = -0.009$, $p = 0.910$, $n = 153$) that this hypothesis seems to lack support, but we cannot test the counter-factual of SB1LP* values that Study 1 respondents would have provided if they had received the contextual information immediately.

5.2.3 | Explaining SB1LP* values

Correlational analyses probed the association of three-trimmed SB1LP* values with demographic (gender, age, education, income, wealth), objective numeracy, and political (trust in agencies to make these regulatory tradeoffs; belief in agency estimate accuracy). For CF ($n = 117$), age ($r = 0.158$), and numeracy ($r = -0.072$), and for LF ($n = 102$) education ($r = -0.121$) and income ($r = -0.120$), followed by age ($r = 0.060$), were the strongest correlations; none were significant at $p < 0.05$. Thus again we see only tiny effects, though constrained by the small sample sizes.

5.2.4 | Explaining "implausible" values

Our third research question was why many lay responses to macro-risk tradeoffs yield implausible imputed SB1LP* values. Table 7 reports descriptive statistics for implausible yielding respondents about their open-ended or multiple-choice answers, or lack of answers, to explain their numerical responses. It shows that low (< $10,000) dominated high (> $1 billion) implausible values, with LF doubling CF's proportion of low responses. A third (32 of 105) of implausible responses came from people failing one or both attention screeners; 18% of the full sample passed both attention checks but offered implausible imputed SB1LP*s. About half of implausibles offered no explanation, including half of those failing attention checks. Thus, inattention is a substantial but incomplete explanation. Some no-explanation respondents probably were "didn't understand," "didn't think carefully," and "don't care" respondents who did *not* select those potentially stigmatizing multiple-choice options. Regulatory beliefs (being pro- or antiregulation generally or for this specific hypothetical regulation) were chosen by 11% of implausibles. A few more (14%) with low imputed values chose the explanation that we should not spend too much on any one cause of death. About 25% chose "preciousness of life" and/or "we should take any opportunity to spend to prevent death." These explanations fit with high-implausible values, but two-thirds came from low-implausibles, implying that people holding these principles did not fully understand how their response might misrepresent this value to policymakers.

Most open-ended responses (47% of implausibles) echoed multiple-choice responses, with some new concepts. For example, LF concerns about prolonging only 100 lives (i.e., they felt prolonging more lives than 100 would better justify offering high cost numbers) may reflect one or more mental processes. These include (1) inconsistency in the logic of how people treated costs versus lives,[7] (2) contrary to the macro-risk premise of a potential shared purpose in prolonging multiple lives, there may be an obverse "anti-shared purpose" (reminding people that a small risk to them is the same as a small number of lives scattered everywhere may make them think the benefit is "lost in the noise"); (3) confusion of regulatory costs with government budgets, despite our earlier extended definition (a confusion observed by Johnson & Finkel, 2016); (4) presuming an improbable short-term interagency budget shift of "excess" regulatory costs to other life-prolonging programs; and/or (5) "pseudo-inefficacy" in valuing human lives (Dickert et al., 2012)—if we cannot save everybody, it is not worth saving fewer people—a view that can sharply restrict charitable donations as the number of people to be helped goes up (e.g., even from one to two victims).

6 | DISCUSSION

6.1 | Findings

We reviewed four sets of data: (1) a reanalysis of our earlier (Finkel & Johnson, 2018) conclusions using our own currently suggested trim and separating data for the two regulations that the earlier article's analysis had combined; (2–3) the two new studies reported fully here; and (4) the

[7] We did not notify people that if you think 100 lives prolonged is too few to worry about, relatively small national costs—divided by 100 million households—may also be too small to worry about. For example, if 330 million Americans each paid a dollar a year to save 100 lives annually, this would correspond to the low end of U.S. agency VSLs, and a third of the highest such VSL.

TABLE 7 Study 3, distribution of implausible responses and multiple-choice explanations

	LF low	CF low	LF high	CF high
Number of implausible responses	65 (16%)	30 (7%)	4 (1%)	6 (1%)
Attention (failed one or both)	17 (26%)	13 (43%)	1 (25%)	1 (17%)
No explanation, neither open-ended or multiple choice	39 (60%)	10 (33%)	0	3 (50%)
Multiple-choice explanations				
Anti- or pro-regulation generally	2 (3%)	2 (7%)	3 (75%)	0
Oppose/support this regulation	2 (3%)	0	2 (50%)	1 (17%)
Many causes of death	7 (11%)	8 (27%)	0	0
Life is precious	10 (15%)	7 (23%)	3 (75%)	1 (17%)
Should spend money if you can prevent death	0	0	3 (75%)	2 (33%)
Didn't understand	5 (8%)	2 (7%)	0	1 (17%)
Didn't think carefully	5 (8%)	0	0	0
Don't care	0	1 (3%)	0	0

Note: Percentages are of $n = 402$ in Number row, and of raw Number value in that column for other rows. People could choose more than one multiple-choice explanation.

study summarized in Supporting Information (despite its limitations, it yielded several results consistent with the other studies)—to test three hypotheses derived from our earlier findings (Finkel & Johnson, 2018) and three research questions about a national-level stated preference approach to estimating the social benefit of one life prolonged (here labeled SB1LP*).

6.1.1 | H1: LF versus CF means

Our first hypothesis—that SB1LP* means would be lower for the LF than CF frame— in our reanalysis of the Finkel and Johnson (2018) data was consistent with data for the EPA regulation, but inconsistent with data for the NHTSA regulation, which parallels the scenario we used in the new studies. It was inconsistent with Study 1 and Study 2 results, but consistent with the separate-information study (Supporting Information). While variation in online panel attributes[8] might explain such mixed results, this seems unlikely given their convergence on the other two hypotheses. Instead, the first macro-risk study (Finkel & Johnson, 2018) on which this hypothesis was based likely happened to yield only one of multiple patterns of comparative means whenever one elicits *unfettered* answers (i.e., no researcher-defined bids) to SB1LP* questions. Lacking any reason to consider either of the CF and LF frames as invalid or superior, perhaps presenting both frames' results and the pooled results would provide a broader set of trimmed means to consider in policy decisions.

[8] The datasets all came from U.S. nationally diverse but nonrepresentative opportunity samples from online panels, but varied in timing (data collected in 2012 [Johnson & Finkel, 2016] versus 2018–2019 for the studies reported here), panel used (a research institute's own carefully curated versus for-profit panels), and reimbursement type (e.g., points traded for rewards versus cash).

6.1.2 | H2: SB1LP* trimmed means versus agency VSLs

Our second hypothesis—that trimmed macro-risk means will be close to or within the range of contemporary US agency-used values of a statistical life derived from microrisk stated or revealed preference studies—was consistent with results from all four datasets.

We note the variation in SB1LP* 2-trimmed means in Table 2 (to allow comparisons with Finkel & Johnson, 2018, and Supporting Information, results). Rounding off, we see trimmed means from $2 million (CF pre-information) to $22 million (LF Study 2), although most (six of 10) are within the $3–10 million range of agency VSLs. Pooling the (three-trim) CF and LF results for post-information values in Study 1, for example, yields a mean of $7,123,387, which compares to a preinformation mean of $1,656,361, or a more than fourfold increase. Given the diversity of values produced by thousands of micro-risk studies, we should find the macro-risk variations neither surprising nor problematic, and the current convergence with agency VSLs might or might not continue as more macro-risk studies are conducted.

One question we urge stated-preference researchers to consider is the validity and goal of any comparison of macro-risk results to agency-used VSLs. There is no reason to take agency VSLs derived from micro-risk studies as a gold standard for benefits valuations. Aside from micro-risk studies' conceptual (e.g., see example in Section 2.1) and methodological issues (e.g., apparently low attention to whether laypeople grasp small probabilities [Section 2.3]; low transparency on trimming methods and central points of VSL distributions [Finkel & Johnson, 2018]), agencies may choose their VSLs to fit their divergent regulatory missions and resources. Further, since national-level tradeoffs inherently free up subjects to express considerations of altruism and "shared purpose" in their responses, while micro-risk tradeoffs are designed to exclude altruism, we would expect imputed SB1LP* values to be higher. The possibly greater intersubject variability seen in the macro-risk results here versus micro-risk results is likely due to the open bid method we used, along with the fact that respondents were rarely confronted directly with tradeoffs that impinged on their personal ability to pay.[9] The micro-risk approach of generally relying on researcher-defined bids and its focus on payment for personal risk reduction, plus the opacity of what trims might have been made on the results, make it unclear whether the comparison with macro-risk results is truly commensurable. If we could compare example(s) of the two kinds of approaches (micro and macro) using similar methods and reporting, those concerned with improving benefits valuation might be better served. On the other hand, assuming that both imputed VSLs and imputed SB1LP* values aim to measure much the same thing (SBNLP, according to Finkel & Johnson, 2018) even if they both fall short in different ways, there should be some convergence, as shown in available macro-risk data. What researchers do not know yet is whether any convergence or gap between micro-risk and macro-risk results is due to unmeasured double-counting of nonpaternalistic altruism (e.g., Johansson, 1992) in macro-risk studies, an issue that we are addressing in a separate article, and/or other factors in either method's results.

Another issue raised by these results is that the pooled (CF + LF) result in Study 2 was $14 million, twice the post-information pooled result of $7 million in Study 1: which version ought regulators to use in decision making, if either? While we cannot explain the difference other than simple variability due to our open-bid process (as noted earlier, we cannot reject the alternative explanation of anchoring on lower preinformation values in Study 1), the higher value is more consistent with the notion that micro-risk and macro-risk methods may be eliciting similarly accurate pictures of public preferences for the benefits of prolonging lives, but that the macro-risk method may also be capturing altruism values, thus leading to a mean higher than the agency VSLs based on only micro-risk methods. This speculation is tentative pending future research directly on whether and how much altruism is captured in the macro-risk approach.

6.1.3 | H3: Implausible values across frames

Our third hypothesis—that implausible values (defined here as < $10,000 and > $1 billion) would be more common for the LF frame—was consistent with results from all four datasets. We had speculated (Finkel & Johnson, 2018) that the LF frame used both here and in all micro-risk studies forces people to think about lives in dollar terms, but in the national scenario it might heighten concern about costs to the economy and/or to the average household, thus lowering imputed SB1LP* values. Consider that $9 per household to reduce a 1 in 1,000,000 personal risk may seem far smaller than spending $900 million collectively (the same $9 per household) to prolong 100 lives nationally, despite our contextual information, and in our open bid process people are free to act upon that feeling of discomfort in a way that most micro-risk studies forestall. Ignoring Study 1 comments on contextual information,[10] the LF frame was much more likely than the CF frame to evoke comments about household affordability (10 comments versus 1 per frame, respectively) and that too few lives were being saved for the regulation to be worthwhile (7, 2), while these frames split comments about the preciousness of life (3, 3). Despite the tiny proportion of such comments in the total of 134, along with the Study 2 results on insufficient lives-prolonged and household costs they imply that cost-consciousness may indeed spur implausibly low values in the LF frame and/or that the CF frame may foster a "spent already" effect that does not require many lives to be saved to justify it (thus reducing the CF proportion of implausibly low SB1LP* values), but this requires more study.

6.1.4 | RQ1: SB1LP* predictors

Effect sizes were small at best and variable across studies regarding which demographic or other factors had the strongest associations with higher versus lower imputed SB1LP* values, thus yielding results from Studies 1–2 (also see Supporting Information) largely consistent with the findings of the earlier macro-risk study (Finkel & Johnson, 2018). It would be premature, however, to conclude that variation in imputed values is not shaped by demographic, numeracy, or political variables, given particularly small samples following our trims. Much larger samples are needed to confirm or refute that hypothesis. However, those micro-risk studies that have examined (mostly demographic) predictors appear to also find relatively weak effects, ones that are inconsistent across studies. If that impression and our own findings on this can be confirmed, it might be that dealing with small probabilities of one's own death and with large amounts of money to prevent 100 national deaths both entail "universal" responses whose variations are not consistently dependent upon demographic or political factors. This suggests that scholars might usefully continue to generalize the central tendency of VSL and SB1LP* distributions across persons regardless of their subgroup, pending

[9] We always told our respondents to assume that costs would be equally distributed across households. The scale of our tradeoffs is roughly $10 per US household per year ($1 billion regulatory costs nationwide over ∼ 100 million households), a bit higher for the rare respondent who chose to spend more than that to prolong 100 lives per year. In a micro-risk task where subjects are asked about a 1/10,000 lifetime risk reduction, dividing a $10 million VSL by 10,000 would entail them contemplating a household "expense" of $1,000, which seems more of a strain on finances than ours. Neither method currently makes these comparisons transparent; doing so might avoid one bias (by providing context) at the risk of another (overemphasizing ability to pay).

[10] We ignore those open-ended comments here as people did not explicitly identify these as decision criteria, nor can they be easily distinguished from similar sounding remarks. They may also reflect cost-consciousness (e.g., CF "the $1 billion started to seem like more of a good deal") or life's value (e.g., LF "The extra information . . . really made me think of the value of life") as well as use of the information. Pursuing this question with closed-ended questions is a more productive research method for this topic, although such questions have their own drawbacks and pitfalls.

TABLE 8 Observations and implications for macro-risk practice and micro-risk introspection

Attributes	Observations	Implications for macro-risk practice	Questions raised for micro-risk practice
Framing	Frame choice did not reliably shape the relative magnitude of imputed SB1LP* means. However, the costs-first (CF) frame poses a more salient question for national regulatory decision-making than the lives-first (LF) frame (Table 1), and it yielded fewer implausible values.	Macro-risk studies should persist with both frames, to further test these findings. But eventually it may be best to focus on CF, if its apparent conceptual and methodological advantages persist. However, policy decisions might benefit from seeing both perspectives, or pooling them, for a more diverse evidential base.	Is micro-risk's LF-only frame to yield the VSL as suited to regulatory decisions as the CF question about benefits that justify spending? We doubt a CF frame will be added, given conceptual difficulty (e.g., "how much personal risk reduction would you require to make its cost of $X worthwhile?"; Finkel & Johnson, 2018, p. 472).
Contextual information	This mix of higher and lower costs and benefits than those for the hypothetical regulation, and perspective clauses, was useful to respondents, raised imputed SB1LP* values far more than it reduced them, and reduced the frequency of implausible values.	Better calibration due to contextual information justifies its continued use. Further qualitative and experimental testing of alternative content and presentation is warranted, plus adding caveats to further reduce frequency of implausible values.	Should micro-risk studies test similar contextual information? Does its apparent absence imply that resulting VSLs might be biased low?
Bounds on bids	The macro-risk approach has elicited respondent-defined bounds between definite support or opposition to the hypothetical regulation, and being unsure; Study 1 added bounds between being unsure and "really unsure," which made little difference to results. Giving people the chance to amend their bounds also made little difference.	Study 1′s results imply that eliciting two bounds, and offering subjects a chance to accept or revise the GM of those bounds, may be superior to presuming there exists a single tradeoff point where support becomes opposition to a regulation. Yet more research would be useful on the impacts of alternative types or numbers of bounds, and why people choose these bounds (e.g., true uncertainty? confidence? better ability to discriminate among numbers of lives worthwhile than among costs that are worthwhile?).	Micro-risk studies do not explicitly ask for the respondent's bounds, but could do so to grasp the respondent's uncertainty (e.g., "what is the largest amount you are sure you would pay to reduce your risk by *X*?" and "the smallest amount you would not pay?"). Interval estimates derived from dichotomous-choice methods are not in fact bounds, but rather researcher imputations from the respondent's responses to researcher-offered bids.
Researcher constraint on bids	The absence of bids used here probably exacerbated the degree of implausible values observed. Macro-risk researchers might constrain lay input to limit implausible values upfront rather than during trimming, but this practice would have its own drawbacks.	Despite some negative respondent reactions—for example, "survey requires too much thinking and too much math!" CF, Study 1—our results indicate that open bidding for a *national*-level tradeoff evokes "plausible" estimates for most people, and only a minority of respondents reported being so mentally challenged by the numbers and tradeoffs that they explicitly or implicitly rejected the results. However, future studies might test screening out people who feel unable to make life-money tradeoffs for moral or political reasons.	Open bids were common in micro-risk studies before the mid-1990s (Cropper et al., 2011), and the practice is still used occasionally (see text, footnote 2). Fixed bids were eventually preferred, in part to avoid socially desirable or strategic answers (e.g., accept a higher-than-acceptable bid in order to register a vote, knowing the higher amount will not be collected). However, another major factor was belief that the bid accept/reject decision was easier (Pearce & Özdemiroglu, 2002, p. 33), which macro-risk results suggest may not be the case. Further, the fixed-bid approach depends in part upon assumptions about the validity of values outside the range of researcher-defined bids (e.g., how one would react to bids higher or lower than those given by the researcher). At minimum, micro-risk investigators should ask whether participants grasp their questions, such as by also seeking useful qualitative or non-bid quantitative responses (e.g., "I would have accepted 100× that highest bid if you'd asked me")

(Continues)

TABLE 8 (Continued)

Attributes	Observations	Implications for macro-risk practice	Questions raised for micro-risk practice
		Both macro/micro-risk: *Do we disenfranchise people who give apparently bizarre answers by trimming, or should we seek to include their perspectives by working harder to understand them? Does excluding "implausible" values merely force upon respondents the researchers' biases about SB1LP*'s magnitude, either our own or those dictated by dozens of prior studies concluding what magnitude of VSL is "plausible"? If our method remedies some deficiencies in prior VSL studies, perhaps what is "plausible" should not depend on what those studies have dictated.*	
Imputation of SB1LP*	Most people chose to "accept" the geometric mean of their bounds as a reasonable tradeoff.	Pending discussion, the GM imputation and "plausible" criteria are worth retaining. Checking the GM's acceptance by respondents is a control on researcher arrogance, but ignores social desirability or fatigue constraints.	As noted earlier, its focus on valuations of risk reductions for oneself raise the question of the method's salience for national regulatory decisions. Should its salience be re-evaluated? Can it be improved?
		Both macro/micro-risk: *Many practical macro-risk and micro-risk problems might be resolved if the aim of valuing the social benefits of prolonging a single human life were made explicit before the tradeoff task, to remove misunderstanding of the task and reduce motives for such things as anti-regulatory protest votes or focusing on per household affordability. Yet resistance to valuing life explicitly may offset benefits of explicitness. Have the benefits and costs of alternative ways to be explicit been fully explored?*	

clarifications (e.g., standardizing trims in VSL studies; removing any double-counting of altruism that might occur in macro-risk estimates).

6.1.5 | RQ2: Contextual information effects

Our second research question—does contextual information about other health and safety costs and lives prolonged affect responses to these national tradeoffs?—was answered in the affirmative: a strong effect in Study 1 in both quantitative (a fourfold increase in the post-information versus pre-information pooled data) and qualitative data, exemplified by a reduction in implausible values, particularly for the LF frame, and an overall increase in values at the individual level in the separate-information study. We cannot compare this degree of learning (about 15% decrease in implausible values in both CF and LF frames in Study 1) to any learning about small probabilities in micro-risk studies, because most if not all of such studies (even if they use the kinds of aids cited earlier) do not conduct pre-/post- within-person or between-person comparisons to see what difference they make. We will share the contextual information before posing any tradeoff questions with all of our future macro-risk subjects (as in Study 2), so we will not be comparing effects of no versus full information in the foreseeable future.

6.1.6 | RQ3: Implausible SB1LP* predictors

Our final research question—what factors contribute to implausible SB1LP* values?—seems to be answered as due to a combination of inattention, regulatory preferences, values, and misinterpretations of how to implement one's values (e.g., on preciousness of life). Micro-risk studies have mostly similar challenges (e.g., Cropper et al., 2011). Ways to attenuate these problems may include retaining attention checks; screening out people who expect to offer protest bids due to antiregulation or life cannot be monetized views; and offering further instructions on how to consider implications of the tradeoff. For example, not all CF respondents might have grasped that in the CF frame one must increase the implied value of life by counter-intuitively *decreasing* the number of lives prolonged that justify the regulation; if not, this might have increased the number of low-magnitude implausible values. We did not test effects of the honesty oath (tested by Jacquemet et al., 2013).

6.2 | Implications for macro-risk versus micro-risk methods

The aim of macro-risk research to date has been to raise questions about how people value the social benefits of life-prolonging and other environmental, health and safety benefits, and perhaps to further develop the approach so that it can complement micro-risk estimates. Excluding the tradeoff perspective itself (national versus personal), any differences in results might be accommodated by adapting either approach's methods for the other. Rather than urge that, however, we follow our earlier example (Johnson & Finkel, 2016) by producing Table 8 to illustrate how certain of our observations might be applied to future macro-risk practice, and might inform introspection by micro-risk scholars about their own methods. The implications for practice listed there do not exhaust possible future research ideas—for example, if the

CF and LF frames yield different results, what happens when you expose a person to both tradeoff scenarios in turn (subject of a forthcoming manuscript)? How does altruism affect imputed SB1LP* results?[11]—but are sufficient to prompt, we hope, much fruitful discussion. We include in Table 8 a couple of questions that apply to both macro-risk and micro-risk methods.

6.3 | Limitations

Although this article increased the macro-risk literature substantially from one to four data collection efforts, it has limits. We used US online opportunity samples only, so we cannot necessarily generalize to adult Americans overall or to non-Americans. However, if challenges here (e.g., implausible numbers; understanding instructions) are due to such issues as skills in cognition, literacy, and/or objective numeracy, making the sample more representative might reduce mean skills and thus exacerbate rather than resolve such problems. We cannot cite our tentative findings on SB1LP* predictors as reassurance that results reflect universality (i.e., little variance across demographics, etc. in SB1LP* values), because much larger samples would be needed to address that more definitively while controlling for potentially large proportions of implausible imputed SB1LP* values. Information content could be added or substituted (e.g., our general-population cost example of gasoline prices was not complemented by a general-population benefit [life-prolonging] example, although the hypothetical regulation is itself an example), or evaluated differently (e.g., our pretest did not elicit any comments about our examples provoking specific values or emotions, but we did not explicitly ask participants about those specific issues). As noted above and in Table 8, our focus here on providing information to aid calibration is only one of several issues around eliciting national tradeoffs.

7 | CONCLUSIONS

Cost-benefit analyses specifically, and regulatory decision-making generally, have benefited—despite considerable debate and criticism—from both revealed and stated preference studies focusing on individuals. We echo arguments that exploring national-level stated preference methods can illuminate strengths and limitations of micro-risk studies (Finkel & Johnson, 2018), but also note that this is part of a larger discussion of valuing public goods (e.g., Bergstrom et al., 2004; Jones-Lee, Hammerton, & Philips, 1985; Koford, 2010). That said, more experiments manipulating the macro-risk method are needed to further probe its strengths and limitations. Our goal is not to determine which method might be superior, or to replace one with the other: we suspect they are better as complementary stated-preference methods, just as micro-risk stated-preference results complement revealed preference results. Ultimately, we hope to encourage broader thinking about how to improve methods of collecting and evaluating information for the benefits side of the benefit-cost analysis ledger.

[11] We did not ask any direct questions that could reveal what proportion of the imputed SB1LP* represented benefits to others versus benefits to self. Measuring altruism in these judgments is on our agenda for a future article.

ACKNOWLEDGMENTS

We appreciate thoughtful comments from three anonymous reviewers. This article was supported by the US National Science Foundation, Decision, Risk, and Management Sciences Program, under Grant No. 1629287. Data collection was overseen by Marcus Mayorga.

ORCID

Branden B. Johnson https://orcid.org/0000-0003-2264-5419

Adam M. Finkel https://orcid.org/0000-0001-6259-9387

REFERENCES

Alberini, A. (2019). Revealed versus stated preferences: What have we learned about valuation and behavior? *Review of Environmental Economics and Policy*, *13*(2), 283–298.

Alolayan, M. A. (2012). (2012). *PM2.5 in Kuwait: Sources, valuation of mortality and benefits of control*. Sc.D. Dissertation, Harvard School of Public Health, p. 247 pp.

Andersson, H. (2006). Willingness to pay for road safety and estimates of the risk of death: Evidence from a Swedish contingent valuation study. *Accident Analysis and Prevention*, *39*, 853–865.

Baik, S., Davis, A. L., & Morgan, M. G. (2019). Illustration of a method to incorporate preference uncertainty in benefit-cost analysis. *Risk Analysis*, *39*, 2359–2368.

Barrio, P. J., Goldstein, D. G., & Hofman, J. M. (2015). *Improving comprehension of numbers in the news*. ACM Conference on Human Factors in Computing Systems (CHI '16). http://cj2015.brown.columbia.edu/papers.html

Bergstrom, J. C., Boyle, K. J., & Yabe, M. (2004). Trading taxes vs. paying taxes to value and finance public environmental goods. *Environmental and Resource Economics*, *28*, 533–549.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, *58*, 739–753.

Braathen, N. A., Lindhjem, H., & Navrud, S. (2011). Valuing mortality risk reductions from environmental, transport, and health policies: A global meta-analysis of stated preference studies. *Risk Analysis*, *31*(9), 1381–1407.

Carson, R. T., & Mitchell, R. C. (2000). Public preferences toward environmental risks: The case of trihalomethanes. In A. Alberini, D. Bjornstad, J., Kahn (Eds.), *Handbook of contingent valuation*. Edward Elgar.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cropper, M., Hammitt, J. K., & Robinson, L. A. (2011). Valuing mortality risk reductions: Progress and challenges. *Annual Review of Health Economics*, *3*, 313–336.

Dickert, S., Västfjäll, D., Kleber, J., & Slovic, P. (2012). Valuations of human lives: Normative expectations and psychological mechanisms of (ir)rationality. *Synthese*, *189*, 95–105.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*, 672–680.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538.

Finkel, A. M., & Johnson, B. B. (2018). The limits of self-interest: Results from a novel stated-preference survey to estimate the social benefits of life-prolonging regulations. *Environmental Law*, *48*, 453–476.

Glickman, M. E., Rao, S. R., & Schultz, M. R. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology*, *67*, 850–857.

Hammitt, J. K., & Treich, N. (2007). Statistical vs. identified lives in benefit-cost analysis. *Journal of Risk and Uncertainty*, *35*, 45–66.

Jacquemet, N., Joule, R.-V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management*, *65*, 110–132.

Johansson, P.-O. (1992). Altruism in cost-benefit analysis. *Environmental and Resource Economics*, *2*, 605–613.

Johnson, B. B., & Finkel, A. M. (2016). Public perceptions of regulatory costs, their uncertainty and interindividual distribution. *Risk Analysis*, *36*, 1148–1170.

Jones-Lee, M. W., Hammerton, M., & Philips, P. R. (1985). The value of safety: Results of a national sample survey. *The Economic Journal*, *95*, 49–72.

Knoblauch, T. A. K., Stauffacher, M., & Trutnevyte, E. (2017). Communicating low-probability high-consequence risk, uncertainty and expert confidence: Induced seismicity of deep geothermal energy and shale gas. *Risk Analysis*, *38*(4), 694–709.

Koford, B. C. (2010). Public budget choices and private willingness to pay. *Public Budgeting and Finance*, *30*, 47–68.

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83.

Pearce, D., & Özdemiroglu, E. (2002). *Economic valuation with stated preference techniques: Summary guide*. Department for Transport, Local Government and the Regions.

Pew Research Center. (2017). *Public trust in government remains near historic lows as partisan attitudes shift*. Pew Research Center for People & the Press.

Pew Research Center. (2018). *Democrats, Republicans give their parties so-so ratings for standing up for "traditional" positions*. Pew Research Center for People & the Press.

Robinson, L. A., Sullivan, R., & Shogren, J. F. (2021). Do the benefits of COVID-19 policies exceed the costs? Exploring uncertainties in the age-VSL relationship. *Risk Analysis*, *41*(5), 761–770.

U.S. Environmental Protection Agency. (2016). *Valuing mortality risk reductions for policy: A meta-analytic approach*. Prepared by the U.S. Environmental Protection Agency's Office of Policy, National Center for Environmental Economics, for review by the EPA's Science Advisory Board, Environmental Economics Advisory Committee. https://yosemite.epa.gov/sab/SABPRODUCT.NSF/81e39f4c09954fcb85256ead006be86e/0CA9E925C9A702F285257F380050C842/$File/VSL+white+paper_final_020516.pdf

Viscusi, W. K. (2020). Pricing the global health risks of the COVID-19 pandemic. Vanderbilt Law Research Paper No. 20–42. Journal of Risk and Uncertainty, *61*, 101–128 https://ssrn.com/abstract=3670841

Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., & De Vries, N. N. K. (2009). Probability information in risk communication: A review of the research literature. *Risk Analysis*, *29*(2), 267–287.

Weinstein, N. D., Sandman, P. M., & Hallman, W. K. (1994). Testing a visual display to explain small probabilities. *Risk Analysis*, *14*(6), 895–896.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Johnson, B. B., Finkel, A. M. (2022). Information effects on lay tradeoffs between national regulatory costs and benefits. *Risk Analysis*, *42,* 2620–2638. https://doi.org/10.1111/risa.13886